On The Convergence Rates Of Empirical Bayes Rules
For Two-Action Problems: Discrete Case *

by

TaChen Liang
Purdue University

Technical Report #87-8

PURDUE UNIVERSITY

DTIC
ELECTE
MAY 2 6 1987
S D
D

DEPARTMENT OF STATISTICS
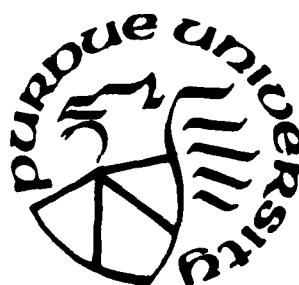
87 5 21 42

On The Convergence Rates Of Empirical Bayes Rules
For Two-Action Problems: Discrete Case *

by

TaChen Liang
Purdue University

Technical Report #87-8

Department of Statistics
Purdue University

March 1987

87-5-21-428

On The Convergence Rates Of Empirical Bayes Rules
For Two-Action Problems: Discrete Case *

by

TaChen Liang

Purdue University

Abstract

The purpose of this paper is to investigate the convergence rates of a sequence of empirical Bayes decision rules for the two-action decision problems where the distributions of the observations belong to a discrete exponential family. It is found that the sequence of the empirical Bayes decision rules under study is asymptotically optimal, and the order of associated convergence rates is $O(\exp(-cn))$, for some positive constant c, where n is the number of accumulated past experience (observations) at hand. Two examples are provided to illustrate the performance of the proposed empirical Bayes decision rules. A comparison is also made between the proposed empirical Bayes rules and some earlier existing empirical Bayes rules.

AMS Subject Classification: 62C12

Key words and phrases: Bayes rule; Empirical Bayes rule; Asymptotically optimal; Rates of convergence

1

1. Introduction

The empirical Bayes approach in statistical decision theory is appropriate when one is confronted repeatedly and independently with the same decision problem. In such instances, it is reasonable to formulate the component problem in the sequence as a Bayes decision problem with respect to an unknown prior distribution on the parameter space and then use the accumulated observations to improve the decision rule at each stage. This approach is due to Robbins (1956, 1964, 1983). Many such empirical Bayes rules have been shown to be asymptotically optimal in the sense that the risk for the *nth* decision problem converges to the optimal Bayes risk which would have been obtained if the prior distribution was fully known and the Bayes rule with respect to this prior distribution was used.

The usefulness of empirical Bayes rules in practical applications clearly depends on the convergence rates with which the risks for the successive decision problems approach the optimal Bayes risk. The purpose of this paper is to investigate the convergence rates of a sequence of empirical Bayes rules for two-action decision problems when the distributions of the observations belong to a discrete exponential family.

Let X be a random observation with probability function of the form

(1.1)
$$f(x|\theta) = h(x)\theta^x\beta(\theta), x = 0, 1, 2, \ldots; 0 < \theta < Q,$$

where $h(x) > 0$ for all $x = 0, 1, 2, \ldots$, and where Q may be finite or infinite. The observation X may be thought of as the value of a sufficient statistic based on several iid observations. Consider the following testing: H_0: $\theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, where θ_0 is a known

2

positive constant. For each $i = 0, 1$, let i denote the action deciding in favor of H_i. For the parameter θ and action i, the loss function is defined as:

$$(1.2) \qquad L(\theta, i) = (1 - i)(\theta_0 - \theta) I_{(0, \theta_0)}(\theta) + i(\theta - \theta_0) I_{[\theta_0, Q)}(\theta),$$

where $I_A(\cdot)$ denotes the indicator function of the set A. In (1.2), the first item is the loss due to taking action 0 when $\theta < \theta_0$, and the second item is the loss of taking action 1 when $\theta \geq \theta_0$. It is assumed that θ is the value of a random variable Θ having an unknown prior distribution $G(\theta)$.

For a decision rule d, let $d(x) = P\{\text{accepting } H_0 | X = x\}$. That is, $d(x)$ is the probability of taking action 0 given $X = x$. Let D be the class of all decision rules. For each decision rule d, let $r(G, d)$ denote the associated Bayes risk. Then, $r(G) = \inf_{d \in D} r(G, d)$ is the minimum Bayes risk among the class D.

Based on the statistical model described above, the Bayes risk associated with the decision rule d is:

$$(1.3) \qquad r(G, d) = \sum_{x=0}^{\infty} [\theta_0 - \varphi(x)] d(x) f(x) + C,$$

where

$$(1.4) \qquad \varphi(x) = \frac{h(x) f(x+1)}{h(x+1) f(x)},$$

$$(1.5) \qquad f(x) = \int_0^Q f(x|\theta) dG(\theta),$$

$$(1.6) \qquad C = \sum_{x=0}^{\infty} \int_{\theta_0}^Q (\theta - \theta_0) f(x|\theta) dG(\theta).$$

3

We consider only priors G such that $\int_0^Q \theta dG(\theta) < \infty$ to insure that the risk is always finite.

Note that C is a constant which is independent of the decision rule d. Thus, from (1.3), a Bayes decision rule, say d_G, is clearly given by

$$(1.7) \qquad d_G(x) = \begin{cases} 1 & \text{if } \varphi(x) \geq \theta_0, \\ 0 & \text{otherwise.} \end{cases}$$

Since the prior distribution G is unknown, it is not possible to apply the Bayes rule for the decision problem at hand. In this situation, we use the empirical Bayes approach. We note that Johns and Van Ryzin (1971) have studied the above decision problem via empirical Bayes approach. In this paper, a sequence of empirical Bayes decision rules $\{d_n^*\}$ is proposed for the above described decision problem. The associated asymptotic optimality property is investigated. It is found that the order of the rate of convergence of $\{d_n^*\}$ is $O(\exp(-cn))$ for some positive constant c, where n is the number of accumulated past experience (observations) at hand. Two examples are given to illustrate the performance of the proposed empirical Bayes decision rules. A comparison is also made between the proposed empirical Bayes rules and some earlier existing empirical Bayes rules.

2. The Proposed Empirical Bayes Rules and Its Asymptotic Optimality

For each $j = 1, \ldots,$ let (X_j, Θ_j) be a pair of random variables, where X_j is observable but Θ_j is not observable. Conditional on $\Theta_j = \theta, X_j$ has probability function $f(x|\theta)$. It is assumed that $\Theta_j, j = 1, \ldots,$ are independently distributed with common unknown prior distribution G. Therefore, $(X_j, \Theta_j), j = 1, 2, \ldots,$ are iid. Let $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ denote the n past observations and let $X_{n+1} \equiv X$ denote the current random observation.

4

According to (1.4) and (1.7), an empirical Bayes decision rule, say d_n^*, is proposed as follows. First, for each $x = 0, 1, 2, \ldots$, let

$$(2.1) \qquad f_n(x) = \frac{1}{n} \sum_{j=1}^{n} I_{\{x\}}(X_j) + \delta_n,$$

where δ_n is a positive value such that $\delta_n = o(1)$. Then, let

$$(2.2) \qquad \varphi_n(x) = \frac{h(x) f_n(x+1)}{h(x+1) f_n(x)}.$$

We then define

$$(2.3) \qquad \varphi_n^*(x) = \left[\max_{0 \le y \le x} \varphi_n(y) \right] \wedge Q,$$

where $a \wedge b = \min\{a, b\}$. Finally, the empirical Bayes decision rule d_n^* is defined as:

$$(2.4) \qquad d_n^*(x) = \begin{cases} 1 & \text{if } \varphi_n^*(x) \ge \theta_0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the past data $\underset{\sim}{X}_n$ is implicitly contained in the subscript n.

Definition 2.1. A decision rule d is said to be monotone if for $x, y \ge 0$ with $x \le y$, $d(x) \le d(y)$.

Note that from (2.3), $\varphi_n^*(x)$ is nondecreasing in x. Then, by (2.4), we see that $d_n^*(x)$ is a monotone decision rule.

In the following, the asymptotic optimality of the sequence of the proposed empirical Bayes decision rules $\{d_n^*\}$ will be investigated. The monotonicity of the decision rules $\{d_n^*\}$ will be used to obtain the related asymptotic optimality.

5

Consider an empirical Bayes decision rule $d_n(x)$. Let $r(G, d_n)$ be the Bayes risk associated with the rule d_n. Then,

$$(2.5) \qquad r(G, d_n) = \sum_{x=0}^{\infty} [\theta_o - \varphi(x)] E[d_n(x)] f(x) + C,$$

where the expectation E is taken with respect to X_n. Since $r(G)$ is the minimum Bayes risk, $r(G, d_n) - r(G) \geq 0$ for all n. Thus, the nonnegative difference $r(G, d_n) - r(G)$ is used as a measure of the optimality of the empirical Bayes decision rule d_n.

Definition 2.2. A sequence of empirical Bayes decision rules $\{d_n\}_{n=1}^{\infty}$ is said to be asymptotically optimal at least of order α_n relative to the (unknown) prior distribution G if $r(G, d_n) - r(G) \leq O(\alpha_n)$ as $n \to \infty$, where $\{\alpha_n\}$ is a sequence of positive numbers such that $\lim_{n \to \infty} \alpha_n = 0$.

Now, straightforward computation leads to that $\varphi(x)$ is increasing in x. Thus, we let $A(\theta_0) = \{x | \varphi(x) > \theta_0\}$ and $B(\theta_0) = \{x | \varphi(x) < \theta_0\}$. Define

$$(2.6) \qquad M = \begin{cases} \min A(\theta_0) & \text{if } A(\theta_0) \neq \phi, \\ \infty & \text{if } A(\theta_0) = \phi, \end{cases}$$

$$(2.7) \qquad m = \begin{cases} \max B(\theta_0) & \text{if } B(\theta_0) \neq \phi, \\ -1 & \text{if } B(\theta_0) = \phi, \end{cases}$$

where ϕ denotes the empty set.

By the increasing property of $\varphi(x)$ with respect to the variable x, $m \leq M$; also, $m < M$ if $A(\theta_0) \neq \phi$. Furthermore,

$$(2.8) \qquad x \leq m \text{ iff } \varphi(x) < \theta_0 \text{ and } y \geq M \text{ iff } \varphi(y) > \theta_0.$$

6

The following theorem is our main result.

Theorem 2.1. Let $\{d_n^*\}$ be the sequence of empirical Bayes decision rules defined above.

Suppose that $\theta_0 < Q$. Also, assume that

(a) $\int_0^Q \theta dG(\theta) < \infty$ and

(b) $m < \infty$.

Then, $r(G, d_n^*) - r(G) \leq O(\exp(-cn))$ for some positive constant c.

Proof: Under Assumption (b) and by (2.8), direct computation leads to

(2.9)

$$r(G, d_n^*) - r(G) = \sum_{x=0}^m [\theta_0 - \varphi(x)] P\{\varphi_n^*(x) \geq \theta_0\} f(x) + \sum_{x=M}^\infty [\varphi(x) - \theta_0] P\{\varphi_n^*(x) < \theta_0\} f(x),$$

where $\sum_{x=0}^m \equiv 0$ if $m = -1$.

The nondecreasing property of $\varphi_n^*(x)$ implies

(2.10)
$$\begin{cases} P\{\varphi_n^*(x) \geq \theta_0\} \leq P\{\varphi_n^*(m) \geq \theta_0\} & \text{for all } x \leq m, \\ P\{\varphi_n^*(x) < \theta_0\} \leq P\{\varphi_n^*(M) < \theta_0\} & \text{for all } x \geq M. \end{cases}$$

Combining (2.9) and (2.10), we have

(2.11) $$r(G, d_n^*) - r(G) \leq b_1 P\{\varphi_n^*(m) \geq \theta_0\} + b_2 P\{\varphi_n^*(M) < \theta_0\},$$

where $0 \leq b_1 = \sum_{x=0}^m [\theta_0 - \varphi(x)] f(x) < \infty, 0 \leq b_2 = \sum_{x=M}^\infty [\varphi(x) - \theta_0] f(x) < \infty$, and the finiteness of both b_1 and b_2 is guaranteed since $\int \theta dG(\theta) < \infty$ by Assumption (a).

Therefore, it suffices to consider the asymptotic behaviors of both $P\{\varphi_n^*(m) \geq \theta_0\}$ and $P\{\varphi_n^*(M) < \theta_0\}$.

7

By the definition of $\varphi_n^*(x)$, when $\varphi_n^*(M) < Q$, then $\varphi_n^*(M) \geq \varphi_n(M)$, where $\varphi_n(\cdot)$ is the function defined in (2.2). In view of this fact and by (2.1) and (2.2),

$$P\{\varphi_n^*(M) < \theta_0\} \leq P\{\varphi_n(M) < \theta_0\}$$

$$(2.12) \qquad = P\{\frac{1}{n}\sum_{j=1}^{n} A_j(M) < -t(M,\theta_0) + \Delta(M,\theta_0,n)\},$$

where

$$(2.13) \qquad A_j(x) = h(x)[I_{\{x+1\}}(X_j) - f(x+1)] - \theta_0 h(x+1)[I_{\{x\}}(X_j) - f(x)],$$

$$(2.14) \qquad t(x,\theta_0) = h(x)f(x+1) - \theta_0 h(x+1)f(x),$$

$$(2.15) \qquad \Delta(x,\theta_0,n) = \delta_n[h(x+1)\theta_0 - h(x)].$$

Also, by the definition of $\varphi_n^*(x)$ and (2.1) and (2.2) again,

$$P\{\varphi_n^*(m) \geq \theta_0\}$$

$$= P\{\varphi_n(y) \geq \theta_0 \text{ for some } y = 0, 1, \ldots, m\}$$

$$\leq \sum_{y=0}^{m} P\{\varphi_n(y) \geq \theta_0\}$$

$$(2.16) \qquad = \sum_{y=0}^{m} P\{\frac{1}{n}\sum_{j=1}^{n} A_j(y) \geq -t(y,\theta_0) + \Delta(y,\theta_0,n)\}.$$

Note that $A_j(x), j = 1\ldots n$, are iid; $E[A_j(x)] = 0$, and $a_1(x,\theta_0) \leq A_j(x) \leq a_2(x,\theta_0)$ where $a_1(x,\theta_0) = -h(x)f(x+1) - h(x+1)\theta_0 + h(x+1)\theta_0 f(x)$ and $a_2(x,\theta_0) = h(x) - h(x)f(x+1) + h(x+1)\theta_0 f(x)$. Also, since $\delta_n = o(1)$ and $m < \infty$, there exists some positive integer n_o such that for all $n \geq n_0$, $|\Delta(y,\theta_0,n)| \leq \frac{1}{2}|t(y,\theta_0)|$ hold for all $0 \leq y \leq m$

8

and for $y = M$. Hence, for n being sufficiently large, $-t(M,\theta_0) + \Delta(M,\theta_0,n) < 0$ since $t(M,\theta_0) > 0$; and $-t(y,\theta_0) + \Delta(y,\theta_0,n) > 0$ for $0 \le y \le m$ since $t(y,\theta_0) < 0$ for $0 \le y \le m$.

In view of the above facts and by Theorem 2 of Hoeffding (1963),

$$P\{\frac{1}{n}\sum_{j=1}^{n} A_j(M) < -t(M,\theta_0) + \Delta(M,\theta_0,n)\}$$

$$\le \exp\{-2n[-t(M,\theta_0) + \Delta(M,\theta_0,n)]^2 a_3^{-1}(M,\theta_0)\}$$

$$(2.17) \qquad \le \exp\{-\frac{n}{2}[-t(M,\theta_0)]^2 a_3^{-1}(M,\theta_0)\}$$

and for $0 \le y \le m$,

$$P\{\frac{1}{n}\sum_{j=1}^{n} A_j(y) \ge -t(y,\theta_0) + \Delta(y,\theta_0,n)\}$$

$$\le \exp\{-2n[-t(y,\theta_0) + \Delta(y,\theta_0,n)]^2 a_3^{-1}(y,\theta_0)\}$$

$$(2.18) \qquad \le \exp\{-\frac{n}{2}[-t(y,\theta_0)]^2 a_3^{-1}(y,\theta_0)\},$$

where $a_3(x,\theta_0) = a_2(x,\theta_0) - a_1(x,\theta_0) = h(x) + h(x+1)\theta_0$.

Let

$$(2.19) \qquad c = \frac{1}{2} \min \{t^2(y,\theta_0) a_3^{-1}(y,\theta_0) | 0 \le y \le m \text{ or } y = M\}.$$

It is clear that $c > 0$ since $m < \infty$ from Assumption (b) and $t^2(y,\theta_0) a_3^{-1}(y,\theta_0) > 0$, for all $0 \le y \le m$ and for $y = M$. Then from (2.11), (2.12) (2.16) to (2.19), we have

$$(2.20) \qquad r(G, d_n^*) - r(G) \le b_1 \sum_{y=0}^{m} \exp(-cn) + b_2 \exp(-cn) = O(\exp(-cn)).$$

Hence, the proof of this theorem is completed.

3. Examples and Remark

The following two examples have been considered by Johns and Van Ryzin (1971) and used to illustrate the performance of their proposed empirical Bayes decision rules for the two-action problem. We cite them and use the same to illustrate the performance of the proposed empirical Bayes decision rules $\{d_n^*\}$.

Example 1. (The Geometric Distribution). Suppose that

$$f(x|\theta) = \theta^x(1-\theta), x = 0, 1, 2, \ldots; 0 < \theta < 1;$$

and that the prior distribution has the probability density function $g(\theta)$ where

$$g(\theta) = (\alpha + 1)(1 - \theta)^\alpha, 0 < \theta < 1, \alpha > -1.$$

Then, $h(x) \equiv 1$ and $f(x) = \frac{(\alpha+1)\Gamma(x+1)\Gamma(\alpha+2)}{\Gamma(x+\alpha+3)}$. Thus, $\varphi(x) = \frac{h(x)f(x+1)}{h(x+1)f(x)} = \frac{x+1}{x+\alpha+3}$ which tends to 1 as $x \to \infty$. Taking $0 < \theta_0 < 1$, then, $A(\theta_0) = \{x|\varphi(x) \geq \theta_0\} \neq \phi$. Therefore, $m < M \equiv \min A(\theta_0) < \infty$. Hence, by Theorem 2.1, $r(G, d_n^*) - r(G) \leq O(\exp(-cn))$ for some positive constant c.

Example 2. (The Poisson Distribution). Let

$$f(x|\theta) = e^{-\theta}\theta^x/\Gamma(x+1), x = 0, 1, 2, \ldots; \theta > 0.$$

Letting the prior density function be $g(\theta) = e^{-\theta}, \theta > 0$, we then have $f(x) = \frac{1}{\Gamma(x+1)} \int_0^\infty \theta^x e^{-2\theta} d\theta = (\frac{1}{2})^{x+1}$, and $h(x) = \frac{1}{\Gamma(x+1)}$. Thus, $\varphi(x) = \frac{h(x)f(x+1)}{h(x+1)f(x)} = \frac{x+1}{2}$ which tends to infinity as x tends to infinity. Therefore, for any finite $\theta_0 > 0, m < \infty$. Then by Theorem 2.1, $r(G, d_n^*) - r(G) \leq O(\exp(-cn))$ for some positive constant c.

10

Johns and Van Ryzin (1971) considered several situations about the behavior of the tail probability of the prior probability density function, under which, their proposed empirical Bayes decision rules may achieve the best possible convergence rate $\alpha_n = n^{-1}$. We also apply those conditions to the sequence of the empirical Bayes decision rules $\{d_n^*\}$. We state the result as a corollary without citing the statement of those conditions. The reader is referred to Johns and Van Ryzin (1971) for detail.

Corollary 3.1. Let $\{d_n^*\}$ be the sequence of the empirical Bayes decision rules defined in Section 2. Suppose that $\int_0^Q \theta \, dG(\theta) < \infty$. Then, either under the assumptions in Theorem 3 or under the assumptions in Theorem 4 of Johns and Van Ryzin (1971), we have $r(G, d_n^*) - r(G) \leq O(\exp(-cn))$ for some positive constant c.

Proof: We need only to verify that $A(\theta_0) \neq \phi$ under each assumption. This can be done directly by noting the Lemmas 4, 5 and 6 of Johns and Van Ryzin (1971).

References

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables *J. Amer. Statist. Assoc.* **58**, 13–30.

Johns, M. V. Jr. and Van Ryzin, J. (1971). Convergence rates for empirical Bayes two-action problems I. discrete case. *Ann. Math. Statist.* **42**, 1521–1539.

Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163, University of California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1-20.

Robbins, H. (1983). Some thoughts on empirical Bayes estimation. *Ann. Statist.* **11** 713-723.

BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
Technical Report #87-8	AD-A181356	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
ON THE CONVERGENCE RATES OF EMPIRICAL BAYES RULES FOR TWO-ACTION PROBLEMS: DISCRETE CASE	Technical
	6. PERFORMING ORG. REPORT NUMBER
	Technical Report #87-8

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
TaChen Liang	N00014-84-C-0167

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Purdue University Department of Statistics West Lafayette, IN 47907	

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Office of Naval Research	March 1987
Washington, DC	13. NUMBER OF PAGES
	13

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayes rule; Empirical Bayes rule; Asymptotically optimal; Rates of convergence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of this paper is to investigate the convergence rates of a sequence of empirical Bayes decision rules for the two-action decision problems where the distributions of the observations belong to a discrete exponential family. It is found that the sequence of the empirical Bayes decision rules under study is asymptotically optimal, and the order of associated convergence rates is $O(\exp(-cn))$, for some positive constant c, where n is the number of accumulated past experience (observations) at hand. Two examples are provided to illustrate the performance of the proposed empirical Bayes decision rules. A comparison is also made between the

DD FORM 1473
1 JAN 73

20. proposed empirical Bayes rules and some earlier existing empirical Bayes rules.

END

7-87

DTIC